Cooperative Correlational and Discriminative Ensemble Classifier Learning for Early Dementia Diagnosis Using Morphological Brain Multiplexes

Raeper, Rory; Lisowska, Anna ; Rekik, Islem

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2018.2863657

Publication date:
2018

Document Version
Peer reviewed version

Cooperative Correlational and Discriminative Ensemble Classifier Learning for Early Dementia Diagnosis Using Morphological Brain Multiplexes

Rory Raeper, Anna Lisowska, Islem Rekik, *Member, IEEE,*
and for the Alzheimer's Disease Neuroimaging Initiative

***Abstract*—Dementia alters the brain wiring on different levels. However, these changes might be subtle particularly in patients with early mild cognitive impairment (eMCI). Hence, developing accurate diagnostic techniques for eMCI identification is critical for early intervention to prevent the onset of Alzheimer's Disease (AD). There is a large body of machine-learning based research developed for classifying different brain states (e.g., AD vs MCI) using neuroimaging data. These works can be fundamentally grouped into two categories. The first one uses *correlational* methods, such as canonical correlation analysis (CCA) and its variants, with the aim to identify most correlated features for diagnosis. The second one includes *discriminative* methods, such as feature selection methods and linear discriminative analysis (LDA) and its variants to identify brain features that discriminate between two brain states. However, existing methods examine these correlational and discriminative brain data *independently*, which overlooks the complementary information provided by both techniques, which could prove to be useful in data classification tasks. On the other hand, how early dementia affects cortical brain *connections in morphology* remains largely unexplored. To address these limitations, we propose a cooperative correlational and discriminative ensemble learning framework for eMCI diagnosis that leverages a brain network representation from multiple morphological networks, each derived from the cortical surface. Specifically, we devise 'the shallow convolutional brain multiplex' (SCBM), which encodes both *region-to-region* and *network-to-network* relationships. Then, we represent each individual brain using a set of SCBMs, which are used to train an ensemble of CCA-SVM and LDA-based classifiers, cooperating to output the label for a new testing subject. Overall, our framework outperformed several state-of-the-art methods including independent correlational and discriminative methods.**

***Index Terms*—Morphological Brain Network, Multi-View Brain Data, Canonical Correlation Analysis, Discriminative Methods, Linear Discriminant Analysis, Ensemble Classifier, Brain Multiplex.**

I. Introduction

THE increasing frequency of dementia occurring is an alarming trend that has prompted urgent research with the goal of preventing the development of the disease. Diagnosing dementia in its early stages is a crucial step in preventing the development of the disease into worsened symptoms [1]. Early mild cognitive impairment (eMCI) is an early stage of dementia which affects the brain function and cognition in subtle ways. These can be challenging to identify when mapping brain connections using Magnetic Resonance Imaging (MRI) of a disordered brain [2]. Developing a deeper understanding of how early stages of dementia alter specific brain connections in patients might improve the likelihood of earlier diagnosis and assist in treating patients. Within this scope, several machine learning approaches leveraged multimodal MRI data including resting-state functional MRI (rsfMRI) and diffusion MRI (dMRI) to distinguish between patients with MCI and healthy controls [3]. However, the very early brain states of dementia including eMCI remain mostly overlooked in previous works when compared with later states such as AD and MCI.

Recent machine-learning methods were devised for MCI identification using connectomic brain data [4], [5]. However, existing works mainly used functional brain networks (derived from rsfMRI) and structural networks (derived from dMRI) [6]. These exclude the recent landmark works [7], [8], [9], which devised morphological brain networks (MBN) for mapping morphological 'connections' in the cortex to circumvent the limitations of functional and structural connectomes [10], [11]. Basically, an MBN is generated by measuring the difference in morphology between two cortical regions based on a specific cortical attribute (e.g., sulcal depth). More importantly, [7], [8], [9] proposed to embed multiple brain networks into a multiplex network structure composed of intra-layer and inter-layer networks. Each intra-layer network in the multiplex represents an MBN derived from a specific cortical attribute, whereas an inter-layer network is a network-to-network similarity inserted between two consecutive intra-layers. The integrated inter-layer network is able to capture high-order brain alterations at the morphological level. While [9] used correlational inter-layers in the brain multiplex structure for late dementia diagnosis, [7] proposed convolutional inter-layers produced by convolving two consecutive MBNs (intra-layers) in the multiplex for early dementia stratification. Notably, both multiplex architectures outperformed conventional single-layer and multi-layer brain network representations. Furthermore,

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

R. Raeper and I. Rekik are affiliated with BASIRA Lab, CVIP group, School of Science and Engineering, University of Dundee, Dundee, UK

A. Lisowska is affiliated with Department of Computer Science, University of Warwick, Coventry, UK

Corresponding author: Islem Rekik, www.basira-lab.com

while [9] used a machine learning method that identifies discriminative connectional features for dementia classification, [7] proposed a correlation-based ensemble learning framework, which identifies highly correlated multiplex features. Such approaches disentangle correlational from discriminative approaches, which might limit our understanding of disordered connectional changes in the diseased brain.

As shown in [7], [9], a prevalent practice for improving classification accuracy is to identify features which contain useful information for the classifier training. In particular, reducing redundant features and primarily focusing on features which have been scrutinized proves to be a useful technique when applied to a classification problem, especially with the growing dimensionality and complexity of data. These existing sample classification approaches can be categorized into two primary groups: methods that aim to identify highly correlated features, and methods that seek to identify the most discriminative features.

The first group, correlational methods, aims to identify highly correlated features within the data, selecting a subset of features from the original data with the purpose of removing redundancy which might hinder prediction accuracy. A wide body of correlational methods can be covered with canonical correlation analysis (CCA) [12], [13], [7], [8] and its variants. CCA, broadly speaking, maps input features into a shared space where features are more comparable and hence, their correlation can be maximized. The projected correlational features in the shared space are then fused together, which reduces the dimensionality of the original data. Several CCA variants have been developed including sparse CCA (sCCA) [14] and non-linear kernel CCA (kCCA) [15]. Specifically, with the proliferation of multi-view data, multi-view CCA (MvCCA) [16], [17] and Tensor CCA (TCCA) [18] were designed with the aim of maximizing the correlation between an arbitrary number of views. In addition to these, there are several innovations upon existing CCA variants such as two-stage kernel CCA (TSKCCA) [19] which by implementing L1-regularization allows for a more reliable identification of non-linear correlations, improving upon KCCA and other existing CCA methods.

The second group, discriminative machine learning approaches, aim to maximize the distinction between sets of data allowing for a reduction in dimensionality. There are a number of discriminative approaches, such as Linear Discriminant Analysis (LDA) [20], where the input features are projected onto a space where their disparity and discriminability are maximized. These were previously used to predict Alzheimer's disease progression from structural imaging [21]. Other methods include discriminative feature selection methods such as Mutual Information (MutInf-FS) [22], which prioritizes minimizing the redundancy in data while maximizing the dependency and relevance of features, or Multi-Cluster Feature Selection (MCFS) [23], thereby reducing the data dimensionality while maintaining its structure. Other approaches such as Infinite Feature Selection (Inf-FS) [24], [9] modeled the relationship between sets of features using a graph to identify groups of highly connected discriminative nodes.

However, a fundamental limitation of the above methods and works reviewed in [25] consists in either identifying correlational or discriminative features for stratifying dementia states. This overlooks the complementary information that can be integrated from both correlational and discriminative approaches to further improve the eMCI/NC classification accuracy.

To fill this gap, we propose a cooperative correlational and discriminative ensemble learning framework, which first pairs brain multiplexes, each generated using a different set of MBNs. Each pair of training multiplexes is then communicated to two different blocks of our framework: the first block includes a set of K discriminative classifiers and the second block includes a set of K correlational classifiers. Ultimately, a pair of testing multiplexes will pass through correlational and discriminative classification blocks, thereby outputting $2K$ labels. By aggregating the labels predicted by all blocks for all pairs of multiplexes, we obtain the final label for the target testing subject. In addition to this landmark contribution, we leverage a multi-layer brain network architecture, the shallow convolutional brain multiplex (SCBM) [8], which unlike the deep CBM proposed in [7], is generated using only two MBNs. This avoids creating redundant features when pairing multiplexes prior to passing them forward to classifiers.

II. Proposed Method

In this section, we design a shallow convolutional brain multiplex to encode low-order and high-order brain connections and present our novel cooperative correlational and discriminative ensemble learning framework. **Fig.** 1 illustrates the different steps for (A) shallow convolutional brain multiplex construction from cortical surface, and (B) multi-source SCBM data pairing for training the correlational block comprising a set of CCA-based SVM classifiers and the discriminative block including a set of LDA classifiers. Below we detail the different steps of our eMCI/NC classification task.

A. Single-view Morphological Brain Network (MBN) Construction

For each cortical attribute (e.g., cortical thickness), we construct a single-view network for each subject. Such network comprises a set of nodes (anatomical brain regions of interest –ROIs) and a collection of edges interconnecting the nodes (representing the dissimilarity between the two brain regions in *morphology*). The average value of a cortical attribute was calculated for each anatomical region of interest (ROI). For each cortical attribute, the strength of each network edge connecting two ROIs is then computed as the absolute difference between their average values, thereby quantifying their dissimilarity (**Fig.** 1). The same procedure was followed to obtain the connectivity matrices from different cortical attributes (e.g., sulcal depth, mean curvature).

B. Shallow Convolutional Brain Multiplex (SCBM) Construction

In a generic way, we define a brain multiplex $\mathcal{M}$ using a set of M intra-layers (or MBNs) $\{\mathbf{V}_1, \ldots, \mathbf{V}_M\}$, each

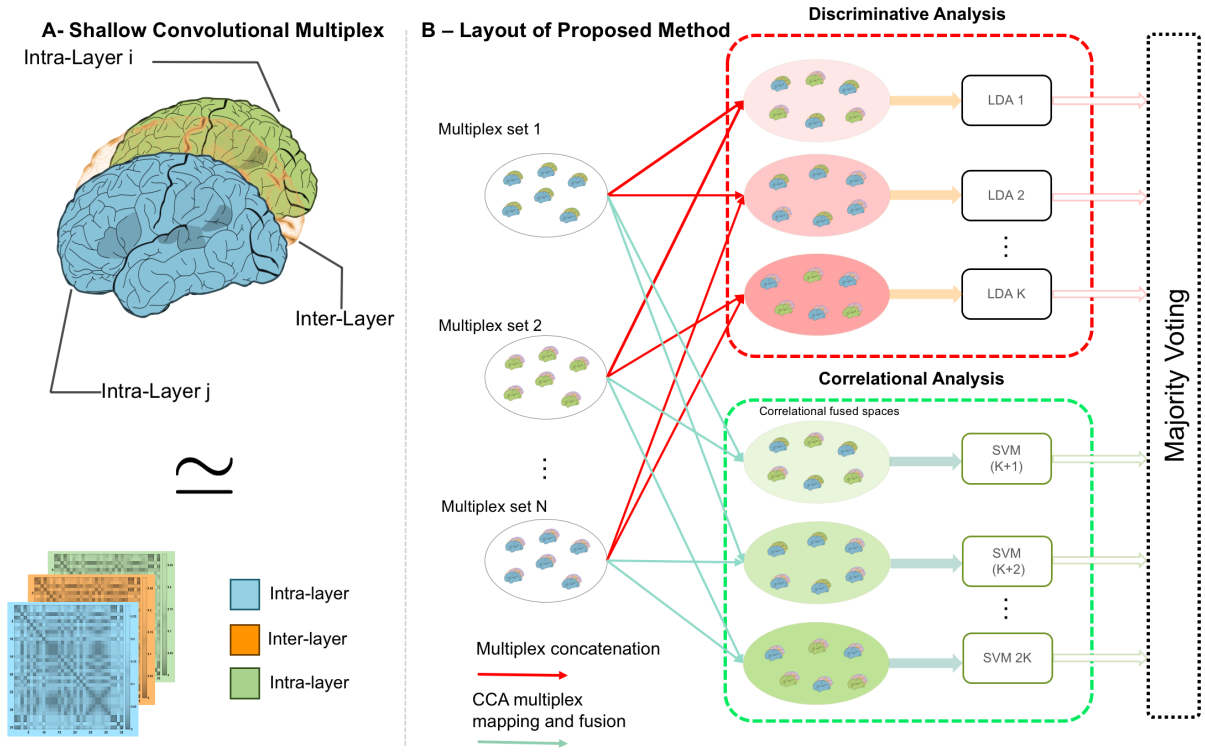

Fig. 1: *Pipeline of the proposed cooperative correlational and discriminative ensemble learning using brain multiplexes.* (A) shows the construction of the multiplex where the inter-layers are created between two intra-layers (two MBNs derived from the cortical surface). (B) For all possible combinations of multiplex pairs, each pair of multiplexes is passed into the ensemble framework, consisting of a correlational learning block (where they are mapped by CCA and classified by SVM) and a discriminative block (where they are mapped and separated into two classes by LDA). The two blocks produce predicted class labels for the test subjects based on analysis of subsequent pairs of multiplexes. The final class label is assigned through majority voting on labels assigned by the two blocks.

representing a single view of the brain morphology (i.e., cortical attribute). Next, we slide an inter-layer $\mathbf{C}_{i,j}$ between two consecutive intra-layers $\mathbf{V}_i$ and $\mathbf{V}_j$. Each inter-layer is created by convolving two consecutive intra-layers. Each element in row a and column b within the convolutional inter-layer matrix $\mathbf{C}_{i,j}$ between views $\mathbf{V}_i$ and $\mathbf{V}_j$ is defined as: $\mathbf{C}_{i,j}(a,b) = \sum_p \sum_q \mathbf{V}_i(p,q)\mathbf{V}_j(a-p+1, b-q+1)$. The multiplex architecture allows not only to explore how different brain views get altered by a specific disorder, but how their relationship might get affected. Since the morphological brain connectivity matrices are symmetric (**Fig.** 1–A), we extract features from each MBN by directly concatenating the off-diagonal weights of all connectivities in each triangular matrix. For each network of size $n \times n$, we extract a feature vector of size $(n \times (n-1)/2)$. Previously, in [7], the generalized multiplex architecture was proposed: $\mathcal{M} = \{\mathbf{V}_1, C_{1,2}, \mathbf{V}_2, \ldots, \mathbf{V}_j, \mathbf{C}_{i,j}, \mathbf{V}_j, \ldots, \mathbf{V}_M\}$. Next, to capture the inter-relationship between all possible combinations of intra-layers in a multiplex, a set of N multiplexes were generated for each subject through reordering the intra-layer networks, thereby generating an *ensemble of brain multiplexes* $\mathbb{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_N\}$. However, this approach resulted in many highly correlated features used for the ensemble learning, which may somewhat mislead classifier learning. To minimize the correlation between different multiplexes when pairing them for ensemble classifier training, we design a shallow (i.e., 2-layer) convolutional brain multiplex structure. We define a *shallow* multiplex $\mathcal{M} = \{\mathbf{V}_i, \mathbf{C}_{i,j}, \mathbf{V}_j\}$ using 2 intra-layers $\mathbf{V}_i$ and $\mathbf{V}_j$ and an inter-layer $\mathbf{C}_{i,j}$ encoding the relationship between $\mathbf{V}_i$ and $\mathbf{V}_j$, slid in between them [8] (**Fig.** 1–A). We note that each subject-specific brain multiplex $\mathcal{M}$ in $\mathbb{M}$ captures unique similarities between 2 different morphological brain network views (e.g., sulcal depth network and cortical thickness network) that are not present in a different shallow multiplex.

C. Proposed Canonical Correlational and Discriminative Mappings of SCBM Sets

Since each multiplex $\mathcal{M}_k \in \mathbb{M}$ captures a unique and complex relationship between different brain network views, one needs to examine all morphological brain multiplexes in the ensemble $\mathbb{M}$. This will provide us with a more holistic understanding of how explicit morphological brain connections can be altered by dementia onset as well as how their implicit high-order (a connection of connections) relationship can be affected. To make use of all the information available from different multiplexes, in the *correlational learning block* of our framework (outlined in green **Fig.** 1–B), we use CCA

[13], [12] to map pairs of multiplex feature vectors extracted from different sets into a shared subspace that depicts highly-correlated relevant features. This correlational block allows to minimize the multiplex set-specific noise and reduces multiplex data dimensionality. Next, we use each fused correlational pair of training multiplex features $\tilde{\mathbf{M}}^c_{k,l}$ to train a linear support vector machine (SVM) classifier (**Fig.** 1–B). Noting that for each training subject we have N multiplexes estimated, we perform C^2_N mappings of each pair of SCBMs in $\mathbb{M}$.

Simultaneously, we train *the paralleled discriminative block* (outlined in red **Fig.** 1–B) aggregating sets of regularized LDA classifiers using the paired SCMBN features from different sets in a *supervised* manner. Additionally, each LDA classifier attempts to maximize the difference between multiplex features so that there are distinct groups based on the given class labels. All training multiplex features are mapped into a discriminative space guided by the labels, where the discriminative paired multiplex features are generated $\tilde{\mathbf{M}}^d_{k,l}$. In the testing stage, we map each pair of testing multiplex feature vector onto their corresponding CCA space where they are communicated to an SVM and LDA classifiers, respectively. Finally, to identify the label of the testing subject, we use majority voting by selecting the highly frequent predicted label outputted by classifiers in both blocks.

III. Results & Discussion

A. Evaluation Data

We evaluated the proposed classification framework using 84 subjects (42 eMCI and 42 NC) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) GO database (adni.loni.usc.edu), each with structural T1-w MR image. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

We used FreeSurfer [26] to reconstruct both right and left cortical surfaces for each subject from T1-w MRI. Then we parcellated each cortical hemisphere into 35 cortical regions using Desikan-Killiany Atlas. For the deep CBM, we defined $N = 6$ multiplexes, each using $M = 4$ MBNs, anchored at $\mathbf{V}_1$. For each cortical attribute (signal on the cortical surface), we compute the strength of the morphological network connection linking i^{th} ROI to the j^{th} ROI as the absolute difference between the averaged attribute values in both ROIs. Multiplex $\mathcal{M}_1$ includes cortical attribute views $\{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4\}$, $\mathcal{M}_2$ includes $\{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_4, \mathbf{V}_3\}$, $\mathcal{M}_3$ includes $\{\mathbf{V}_1, \mathbf{V}_3, \mathbf{V}_4, \mathbf{V}_2\}$, $\mathcal{M}_4$ includes $\{\mathbf{V}_1, \mathbf{V}_3, \mathbf{V}_2, \mathbf{V}_4\}$, $\mathcal{M}_5$ includes $\{\mathbf{V}_1, \mathbf{V}_4, \mathbf{V}_2, \mathbf{V}_3\}$, and $\mathcal{M}_6$ includes $\{\mathbf{V}_1, \mathbf{V}_4, \mathbf{V}_3, \mathbf{V}_2\}$. For each cortical region, $\mathbf{V}_1$ denotes the maximum principal curvature brain view, $\mathbf{V}_2$ denotes the mean cortical thickness brain view, $\mathbf{V}_3$ denotes the mean sulcal depth brain view, and $\mathbf{V}_4$ denotes the mean average curvature brain view. As for the proposed SCBM, we define $N = C^2_4 = 6$ shallow multiplexes, by considering all possible pairings of 2 views out of 4. For our experiments, we created 4 representations of MBN data: (1) 'Views' by concatenating all MBNs, (2) 'Correlational multiplexes' with inter-layer computed using Pearson correlation, (3) 'Convolutional multiplexes' composed of 4 intra-layers with inter-layers generated using 2D convolution, and (4) 'Shallow convolutional multiplexes' composed of 2 intra-layers with inter-layers generated using 2D convolution.

Remark: In the convolutional brain multiplexes, the convolution operation between intra-layers captures the signal within a subgraph (a small patch in the connectivity matrix) extracted from a first layer (whole matrix) as an expression of other subgraphs extracted from a second layer. One can think of the inter-layer network as a 'high-order blending' of both intra-layers, expressing the amount of overlap of intra-layer 1 as it is shifted over intra-layer 2.

B. Experiment Setup

To demonstrate the effectiveness of integrating correlational and discriminative methods into a single framework, we benchmarked our method against several discriminative methods including: Eigenvector Centrality (ECFS) [27], Mutual Information (MutInf-FS) [28], and Infinite Feature Selection (Inf-FS) [24]. We also benchmarked our method against the CCA-based eMCI/NC classification framework in [7]. We also evaluated the performance of each of the aforementioned discriminative methods when combined with CCA using our proposed framework. Additionally, we benchmarked against newer discriminative and correlational methods, Tensor CCA (TCCA) [18], the multi-view Discriminant Analysis (MvDA) [29], and finally the two methods combined into a paired classifier.

The first method, Tensor CCA (TCCA) [18], utilizes tensors for a correlation analysis of an arbitrary number of views. Specifically, TCCA maps multiple views into covariance tensors where correlated features can be identified, maximizing the correlation between several views and hence improving upon traditional CCA methods which are optimized for pairwise correlation. Furthermore, TCCA is capable of identifying high order correlation information by also adopting the alternating least squares (ALS) algorithm, further improving upon its correlational counterparts. The second method, Multi-view Discriminant Analysis (MvDA) [29], follows a similar process by extending traditional LDA to support multiple views. MvDA aims to maximize the between-class variations while simultaneously minimizing any within-class variance, consequently highlighting discriminative features. Subsequently, we combine both MvDA and TCCA into a single framework for a final benchmark, to allow for the analysis of the shared information between the correlational and discriminative methods.

We used leave-one-out cross-validation (LOO) to evaluate our proposed method and its comparison methods. Specifically, using a support vector machine (SVM), we train a model on the data with the number of subjects minus one and subsequently predict a label for the remaining subject. This process is repeated until there are predicted labels for

Table 1. *Average eMCI/NC classification accuracy using our method and different comparison methods.*

		Left Hemisphere			Right Hemisphere		
Method	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ensemble SVM Paired Classifiers using CCA [7]	Views	59.52	52.38	66.67	67.86	61.9	73.81
	Correlational	65.48	69.05	61.9	58.33	64.29	52.38
	Convolutional	64.29	66.67	61.9	71.43	73.81	69.05
	Shallow Conv	63.1	66.67	59.52	73.81	76.19	71.43
Ensemble SVM Paired Classifiers using ECFS [26]	Views	58.33	50	66.67	73.81	64.29	83.33
	Correlational	65.29	61.9	66.67	73.81	73.81	73.81
	Convolutional	63.1	64.29	61.9	76.19	73.81	78.57
	Shallow Conv	70.24	76.19	64.29	66.67	66.67	66.67
Ensemble SVM Paired Classifiers using CCA [11] + ECFS [26]	Views	59.52	52.38	66.67	69.05	64.29	73.81
	Correlational	53.57	57.14	50	57.14	61.9	52.38
	Convolutional	57.14	59.52	54.76	70.24	73.81	66.67
	Shallow Conv	63.1	66.67	59.52	78.57	78.57	78.57
Ensemble SVM Paired Classifiers using MutInf-FS [27]	Views	61.9	50	73.81	72.62	66.67	78.57
	Correlational	55.59	52.38	59.52	63.1	61.9	64.29
	Convolutional	57.14	57.14	57.14	64.29	64.29	64.29
	Shallow Conv	66.67	71.43	61.9	76.19	78.57	73.81
Ensemble SVM Paired Classifiers using CCA [11] + MutInf-FS [27]	Views	64.29	59.52	69.05	66.67	61.9	71.43
	Correlational	64.29	69.05	59.52	54.76	57.14	52.38
	Convolutional	58.33	61.9	54.76	71.43	71.43	71.43
	Shallow Conv	63.09	69.05	57.14	77.38	78.57	76.19
SVM Classifier using MvDA [28]	Views	66.67	66.67	66.67	78.57	85.71	71.43
	Correlational	66.67	66.67	66.67	72.62	71.43	73.81
	Convolutional	69.05	71.43	66.67	76.19	73.81	78.57
	Shallow Conv	69.05	69.05	69.05	78.57	78.57	78.57
Ensemble SVM Paired Classifiers using TCCA [17]	Views	71.43	66.67	76.19	75	71.43	78.57
	Correlational	71.43	76.19	66.67	66.6	69.05	64.29
	Convolutional	69.05	69.05	69.05	70.24	69.05	71.43
	Shallow Conv	69.05	73.91	64.29	75	78.57	71.43
Ensemble SVM Paired Classifiers using MvDA [28] + TCCA [17]	Views	71.43	73.81	69.05	75	90.45	59.52
	Correlational	70.24	85.71	54.76	66.67	69.05	64.29
	Convolutional	70.24	69.05	71.43	73.81	88.1	59.52
	Shallow Conv	72.62	73.81	71.43	77.38	69.05	85.71
Ensemble LDA Paired Classifiers [19]	Views	73.81	76.19	71.43	70.24	61.9	78.57
	Correlational	**76.19**	**78.57**	**73.81**	71.43	71.43	71.43
	Convolutional	73.81	78.57	69.05	77.38	78.57	76.19
	Shallow Conv	73.81	80.95	66.67	73.81	73.81	73.81
Ensemble LDA [19] and CCA-SVM [11] Paired Classifiers **(Ours)**	Views	67.86	76.19	59.52	70.24	69.05	71.43
	Correlational	69.05	66.67	71.43	70.24	66.67	73.81
	Convolutional	70.24	73.81	66.67	**79.76**	**78.57**	**80.95**
	Shallow Conv	72.62	78.57	66.67	**80.95**	**83.33**	**78.57**

every subject where they can be compared to the ground truth labels, producing an accuracy, sensitivity and specificity scores. Furthermore, due to MvDA containing an optimization variable, λ, we tuned it using 5-fold cross-validation, starting with $\lambda = 0.1$ and iterating through to 0.9 with a step size of 0.1. This form of nested cross-validation was used for two other variables in the experiment. Specifically, the number of selected features for ECFS and MutInfFS were automatically tuned using a similar nested 5-fold cross-validation technique, where the feature size varied from 50 to 400 with a step size of 50. Additionally, we empirically tuned specific variables in TCCA due to the cumbersome runtime of the method. We set the optimization variable ϵ to 0.5 and tuned the number of features for each data type. Finally, ECFS, TCCA, and MvDA required dimensionality reduction on high-dimensional data, due to computational limitations and memory overload. Consequently, PCA was applied to the data for dimensionality reduction. In particular, for ECFS method, we applied PCA to

the deep convolutional multiplexes as it was unable to handle high dimensional data.

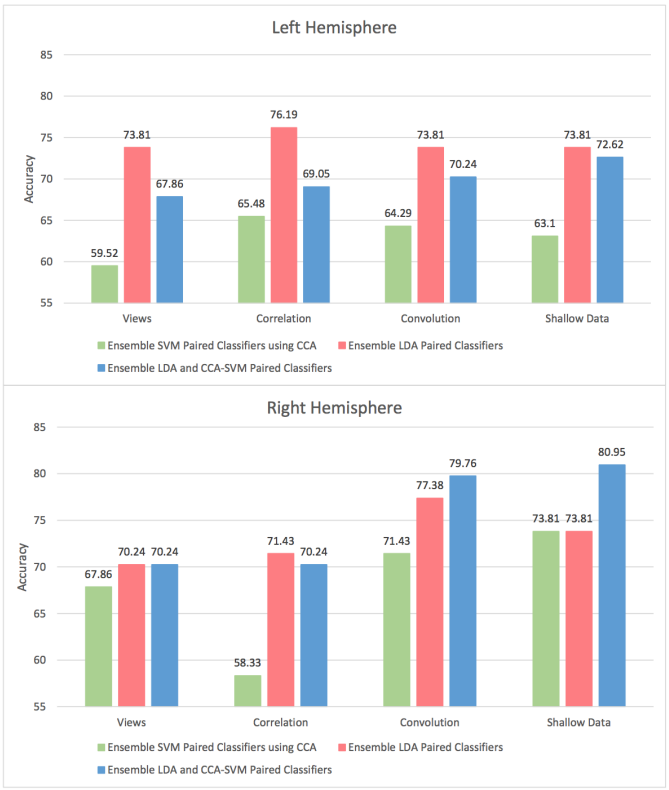

Fig. 2: *Classification accuracy for all data types in both the left and right hemispheres.* Shown are the proposed method, Ensemble LDA and CCA-SVM Paired Classifiers, and its derivative methods Ensemble SVM Paired Classifiers using CCA and Ensemble LDA Paired Classifiers. All methods were applied to different multi-view brain network data representations.

Table 1 displays the accuracy, sensitivity, and specificity for each method. Each benchmark method shows results for morphological views, correlational multiplexes [9], deep convolutional multiplexes [7], and shallow convolutional multiplexes in both hemispheres. Additionally, in **Fig.** 2, we show the comparison between independent LDA, CCA, and our proposed method via column graph, showing the accuracy for each data type.

The best accuracy was attained using the proposed framework with shallow convolutional multiplexes for the right hemisphere, with an accuracy of 80.95% (**Fig.** 2). The proposed framework improved upon the independent methods, increasing the accuracy by 3-7% in the right hemisphere. The left hemisphere results show a contrasting conclusion, achieving the best accuracy using independent LDA with correlational multiplexes (76.19%). These conclusions can be seen clearly in **Fig.** 2. This difference in accuracy between hemispheres might be explained by the way early dementia biomarkers manifest in the data, sometimes being more difficult for certain methods to identify.

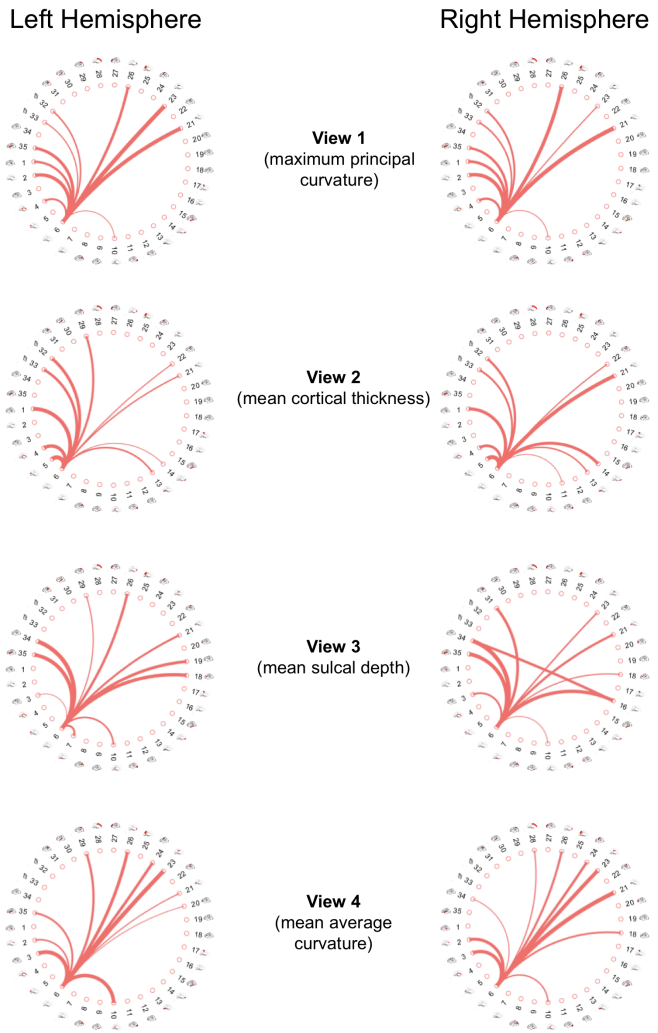

Fig. 3: *Circular graphs displaying the top 10 most discriminative features and the corresponding regions of the brain identified by ECFS.* Each view denotes a certain feature: View 1 shows the maximum principle curvature; View 2 shows the mean cortical thickness; View 3 shows the mean sulcal depth; and View 4 shows the mean average curvature.

C. Identifying the most discriminative features

To identify the top 10 most discriminative features of each view (cortical region), we applied ECFS which returned a ranking of features, from the most to the least discriminative which allowed us to select the optimal number of features. We note that we used ECFS as it is the second best discriminative method that allows to track the original features unlike LDA which projects the features. The ranked features were averaged between all subjects to produce circular graphs showing the most common top features for each view (**Fig.** 3). Weights were then produced for each feature where they were correlated with the thickness of the edge in the circular graph. The most highly correlated features are not shown due to neither TCCA nor CCA returning a ranking of features but directly projecting the data within their methods, which inhibits tracking of the original features.

There are several compelling observations that can be made from the circular graphs shown in **Fig.** 3. The most prominent of which is the significance of the *entorhinal cortex* which appears to be a primary hub in all views and both hemispheres, as it is present in every connection. This also may indicate the significance of this region as a strong biomarker for detecting early stages of dementia. Another notable observation occurs in the right hemisphere's view 3 (mean sulcal depth) where there is a strong connection between the *transverse temporal cortex* and the *parahippocampal gyrus*, neither of which appear in any other view. This may be due to the *parahippocampal gyrus'* proximity to the *entorhinal cortex*, causing it to be affected by early stages of dementia. Finally, the *posterior-cingulate cortex* is a prominent feature in views 1 (maximum principle curvature) and 4 (mean average curvature), occurring in both hemispheres, although more pronounced in the left. Like the *entorhinal cortex*, the *posterior-cingulate cortex* is a highly connected hub node, and may also be a key biomarker to aid in the diagnosis of early dementia.

D. Performance of proposed method

We proposed a novel framework that pairs discriminative and correlational methods together to leverage the complementary information that can be neglected when only one of the approaches is used. Furthermore, we improved upon existing deep convolutional brain networks by using a 2-layer alternative which shows the relationship between 2 views, while also reducing the high dimensionality and redundant features of the deep convolutional multiplexes.

Our proposed framework significantly outperformed several comparison methods when focusing on both the deep and shallow brain multiplexes in the right hemisphere. These conclusions are shown clearly in **Fig.** 2, where the combined LDA and CCA method has a large margin of increase over their independent counterparts. The left hemisphere produced contrasting results, with independent Linear Discriminant Analysis (LDA) achieving the highest accuracy using correlational data. These opposing results might indicate that early stages of dementia affect the hemispheres in distinct ways. While also producing superior results with the proposed method, it was also observed that pairing both correlational and discriminative methods significantly improved results for data in the right hemisphere, when compared to any of the independent discriminative methods. Both ECFS and MutInf-FS improved their accuracy when paired with CCA for the shallow data in the right hemisphere, yielding their highest results, further demonstrating the proposed framework's improvement.

The benchmark methods maintained a consistent performance with the proposed framework; however, the performance in the left hemisphere was notably lower. One interesting result was with the newly proposed methods TCCA, MvDA and their paired classifier, which achieved very high accuracies individually but were somehow weakened when paired. One potential factor for their reduced accuracies was the addition of principle component analysis (PCA) as neither method can directly handle high-dimensional data. PCA has proven to affect the accuracy of classification as it can reduce reduncancy; however, it can also remove useful features and hence affect the performance either way [30], [31]. The effect of PCA may also be noted within the fluctuations in the performance of ECFS using convolutional data.

E. Performance of Left & Right Hemispheres

An interesting observation to note is the disparity between accuracies from the left and right hemispheres. While the right hemisphere performs relatively well, the left severely underperforms with almost all methods, including the proposed framework, with the exception of Linear Discriminant Analysis (LDA). Despite the contrasting conclusion from the right hemisphere, pairing discriminative and correlational methods is detrimental to the classification accuracy in the left hemisphere when compared to standalone discriminative methods. This may be explained by the severe underperformance of the CCA fusion method, which can be seen in **Table** 1, where there is up to a 10% decrease in accuracy between hemispheres. This disparity can be observed in **Fig.** 2 where the classification accuracy significantly improves when the correlational and discriminative accuracies are similar individually. However, when the disparity between the correlational and discriminative methods is large, the accuracy generally averages between the two instead of improving. This is probably due to the majority voting being heavily biased towards the lower accuracy, pushing the classifier to select the incorrect label.

F. Analysis of Most Discriminative Features

Using ECFS to identify the top most discriminative features for each view, we could identify significant biomarkers for displaying early signs of dementia. The most notable finding was the prominence of the entorhinal cortex where almost all the top 10 features were connected. While being a hub node in the brain responsible for numerous connections, the entorhinal cortex is also notable for being one of the first regions to be affected by early states of dementia [32], [33] and is hence a crucial biomarker to analyze when diagnosing eMCI. Furthermore, another notable region was the posterior cingulate cortex which occurs multiple times in the maximum principle curvature and the average mean curvature views. The prominence of this identified region can possibly be explained by the proximity to the affected entorhinal cortex, indicating that the spread of dementia starts at hub nodes then spreads outwards into connected regions. Interestingly, the posterior-cingulate cortex, which also appeared frequently, is another dominant hub node commonly identified as one of the first affected regions by dementia [34]. These findings may assist in identifying early states of dementia by clearly displaying significant biomarkers and patterns in brain scans of the affected patients. Additionally, it was observed that the most prominent identified features remained consistent across all the morphological brain views, which might indicate that dementia affects the brain morphology uniformly. Further analysis could be made for identifying prominent features by finding a correlational method which allows us to better track the most relevant features.

G. Shallow vs. Deep Convolutional Brain Multiplexes

A limitation of the deep convolutional multiplexes was the possible redundancy of data due to its high dimensionality. Since different deep multiplexes contain overlapping sets of features, resulting in highly-correlated input data, it might result in a suboptimal ensemble performance. Hence, the shallow multiplex structure solved this problem by reducing the correlation between individual classifiers in the ensemble and *overall* produced a better ensemble classifier performance compared to the ensemble classifier using deep convolutional multiplex structure. The utilized shallow convolutional brain multiplex (SCBM) consistently outperformed concatenated MBN views and correlational brain multiplexes across all methods within the right hemisphere –except for independent ECFS.

H. Limitations & Future Work

To further improve the proposed classification architecture, it is essential to address the limitations found with the analysis of the results attained in our experiment. In particular, the performance of the proposed framework and comparison methods within the left hemisphere are a significant limitation. Both the correlational and discriminative methods have a large disparity in their performance between hemispheres, with the right hemisphere consistently giving better results. The benchmark CCA method performs particularly poorly in the left hemisphere, and noticeably has a negative effect upon the paired method's performance. Generally, the left hemisphere contains features which might be harder to identify, with dementia possibly manifesting in significantly different ways across hemispheres. Additionally, the discriminative methods perform poorly on the left hemisphere, with both ECFS and MutInf-FS losing a considerable accuracy.

Finding methods which further improve the results of the left hemisphere would be a suitable initial direction for future work. For improving the correlational block of our framework, one can use methods that go beyond linear correlations and better identify sparse feature correlations such as Sparse CCA, or Kernel CCA. Additionally, a more tuned optimization of Tensor CCA (TCCA) may also be applicable to the problem as it showed promising results. Improving the discriminative methods could also have a significant impact on improving the combined accuracy. Finding more discriminative feature selection methods which rank features, like ECFS and MutInf-FS, may be a useful addition as a more accurate method could identify more biomarkers within the left hemisphere. Furthermore, tuning the selected features for ECFS and MutInf-FS, by using smaller step-sizes, could have a positive affect on the classification as they would identify only the most reliable features that differentiate between both groups.

IV. Conclusion

Diagnosing early brain symptoms of dementia such as early Mild Cognitive Impairment (eMCI) is vital to prevent worsening of symptoms. To assist this diagnosis, we proposed a cooperative correlational and discriminative ensemble learning framework using shallow convolutional brain multiplexes. Our method attained a large increase in accuracy when using both the shallow and deep convolutional data against several benchmark methods including [7], and numerous discriminative methods. A reported increase of over 7% was attained for the shallow data which supports our theory that utilizing both correlational and discriminative analysis methods yields an increase in overall performance when focusing on the right hemisphere. Another conclusion drawn from these results is that similar accuracy between the shallow and deep convolutional data is obtained with the shallow having a higher prediction accuracy frequently. This shows that investigating the similarity between two brain networks can be convenient when analyzing the multi-level effects dementia has on brain connections. In our future work, we will explore other correlational and discriminative methods that are based on feature ranking and avoid projections, thereby allowing to identify the original features most relevant to the target classification task. Inspired by how convolutional neural networks work, we will also leverage existing deep learning methods [35] to automatically *learn* multiplex inter-layers for a more powerful and generalizable training of our ensemble.

Acknowledgment

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1] R. E. Amariglio, M. C. Donohue, G. A. Marshall, D. M. Rentz, D. P. Salmon, S. H. Ferris, S. Karantzoulis, P. S. Aisen, and R. A. Sperling, "Tracking early decline in cognitive function in older individuals at risk for Alzheimer disease dementia: the Alzheimers Disease Cooperative

Study Cognitive Function Instrument," *JAMA neurology*, vol. 72, no. 4, pp. 446–454, 2015.
[2] S. W. Moon, I. D. Dinov, S. Hobel, A. Zamanyan, Y. C. Choi, R. Shi, P. M. Thompson, A. W. Toga, and A. D. N. Initiative, "Structural brain changes in early-onset alzheimer's disease subjects using the loni pipeline environment," *Journal of Neuroimaging*, vol. 25, no. 5, pp. 728–737, 2015.
[3] C. J. Brown and G. Hamarneh, "Machine learning on human connectome data from MRI," *arXiv preprint arXiv:1611.08699*, 2016.
[4] X. Chen, H. Zhang, Y. Gao, C.-Y. Wee, G. Li, and D. Shen, "High-order resting-state functional connectivity network for mci classification," *Human brain mapping*, vol. 37, no. 9, pp. 3282–3296, 2016.
[5] C.-Y. Wee, S. Yang, P.-T. Yap, D. Shen, A. D. N. Initiative *et al.*, "Sparse temporally dynamic resting-state functional connectivity networks for early mci identification," *Brain imaging and behavior*, vol. 10, no. 2, pp. 342–356, 2016.
[6] A. Fornito, A. Zalesky, and M. Breakspear, "The connectomics of brain disorders," *Nature Reviews Neuroscience*, vol. 16, no. 3, p. 159, 2015.
[7] A. Lisowska, I. Rekik, A. D. N. Initiative *et al.*, "Pairing-based ensemble classifier learning using convolutional brain multiplexes and multi-view brain networks for early dementia diagnosis," *International Workshop on Connectomics in Neuroimaging*, pp. 42–50, 2017.
[8] ——, "Joint pairing and structured mapping of convolutional brain morphological multiplexes for early dementia diagnosis," *Brain Connectivity*, 2018.
[9] I. Mahjoub, M. A. Mahjoub, and I. Rekik, "Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states," *Nature Scientific Reports*, 2018.
[10] B. Chen, T. Xu, C. Zhou, L. Wang, N. Yang, Z. Wang, H.-M. Dong, Z. Yang, Y.-F. Zang, X.-N. Zuo *et al.*, "Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one month," *PLoS One*, vol. 10, no. 12, p. e0144963, 2015.
[11] D. Petrov, A. Ivanov, J. Faskowitz, B. Gutman, D. Moyer, J. Villalon, N. Jahanshad, and P. Thompson, "Evaluating 35 methods to generate structural connectomes using pairwise classification," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 515–522, 2017.
[12] X. Zhu, H.-I. Suk, S.-W. Lee, and D. Shen, "Canonical feature selection for joint regression and multi-class identification in alzheimer's disease diagnosis," *Brain imaging and behavior*, vol. 10, no. 3, pp. 818–828, 2016.
[13] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
[14] D. R. Hardoon and J. Shawe-Taylor, "Sparse canonical correlation analysis," *Machine Learning*, vol. 83, no. 3, pp. 331–353, 2011.
[15] S. Akaho, "A kernel method for canonical correlation analysis," *arXiv preprint cs/0609071*, 2006.
[16] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," *Conference on Data Mining and Data Warehouses (SiKDD 2010)*, pp. 1–4, 2010.
[17] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE transactions on image processing*, vol. 11, no. 3, pp. 293–305, 2002.
[18] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction," *IEEE transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3111–3124, 2015.
[19] K. Yoshida, J. Yoshimoto, and K. Doya, "Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data," *BMC bioinformatics*, vol. 18, no. 1, p. 108, 2017.
[20] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.*, pp. 41–48, 1999.
[21] S. F. Eskildsen, P. Coupé, V. S. Fonov, J. C. Pruessner, and D. L. Collins, "Structural imaging biomarkers of Alzheimer's disease: predicting disease progression," *Neurobiology of aging*, vol. 36, pp. S23–S31, 2015.
[22] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
[23] D. Cai, C. Zhang, and X. He, "Unsupervised feature selection for multi-cluster data," *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 333–342, 2010.
[24] G. Roffo, S. Melzi, and M. Cristani, "Infinite feature selection," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4202–4210, 2015.
[25] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, 2017.
[26] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.
[27] G. Roffo and S. Melzi, "Features selection via eigenvector centrality," *Proceedings of New Frontiers in Mining Complex Patterns (NFMCP 2016)(Oct 2016)*, 2016.
[28] M. Bennasar, Y. Hicks, and R. Setchi, "Feature selection using joint mutual information maximisation," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8520–8532, 2015.
[29] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 188–194, 2016.
[30] H. Park, J.-j. Yang, J. Seo, and J.-m. Lee, "Dimensionality reduced cortical features and their use in the classification of alzheimer's disease and mild cognitive impairment," *Neuroscience letters*, vol. 529, no. 2, pp. 123–127, 2012.
[31] M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, M. Gómez-Río, A. D. N. Initiative *et al.*, "Principal component analysis-based techniques and supervised classification schemes for the early detection of alzheimer's disease," *Neurocomputing*, vol. 74, no. 8, pp. 1260–1271, 2011.
[32] U. A. Khan, L. Liu, F. A. Provenzano, D. E. Berman, C. P. Profaci, R. Sloan, R. Mayeux, K. E. Duff, and S. A. Small, "Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical alzheimer's disease," *Nature neuroscience*, vol. 17, no. 2, p. 304, 2014.
[33] M. E. López, R. Bruna, S. Aurtenetxe, J. Á. Pineda-Pardo, A. Marcos, J. Arrazola, A. I. Reinoso, P. Montejo, R. Bajo, and F. Maestú, "Alpha-band hypersynchronization in progressive mild cognitive impairment: a magnetoencephalography study," *Journal of Neuroscience*, vol. 34, no. 44, pp. 14 551–14 559, 2014.
[34] R. Leech and D. J. Sharp, "The role of the posterior cingulate cortex in cognition and disease," *Brain*, vol. 137, no. 1, pp. 12–32, 2013.
[35] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 6, pp. 2063–2079, 2018.

Rory Raeper is an undergraduate student, currently studying for a BSc (Hons) in Applied Computing at the University of Dundee. As part of this, in 2017, he started work researching machine learning methods for healthcare. His research interests are primarily focused on fields relating to machine learning and AI.

Anna Lisowska is a PhD student in the Computer Science department at Warwick University. She did her Master's degree in Data Engineering at the University of Dundee after graduating with a BSc diploma in Biomedical Sciences (Neuroscience) from the University of Edinburgh. Currently, she is a member of Tissue Image Analytics Lab at Warwick University, carrying out research in Computational Biology focused on combining analysis of genetic data and medical image data for an improved patient diagnosis, treatment recommendation and outcome prediction.

Islem Rekik is an assistant professor (lecturer) within Computing at the University of Dundee. Previously, she was a postdoctoral research scholar at the IDEA lab (University of North Carolina) following a PhD in "Neuroimaging and Computer Sciences" from the University of Edinburgh in 2014. Currently, she is the director of BASIRA (Brain And SIgnal Research & Analysis,www.basira-lab.com) laboratory and a member of the Computer Vision & Image Processing (CVIP) group at the University of Dundee. She is a program committee member of leading international medical image analysis conferences and workshops including MICCAI, Machine Learning in Medical Imaging (MLMI), Connectomics in NeuroImaging (CNI), and Deep Learning in Medical Imaging (DLMI). She co-organized Predictive Intelligence in MEdicine (PRIME) and CNI workshops, in conjunction with MICCAI 2018. Her research work aims to infuse advanced computer vision and machine-learning methods into big neuroimaging and signal data analysis for improving healthcare and wellbeing.